

Synergistic use of GPCR modelling and SDM experiments to understand ligand binding

Running Header: GPCR modelling and SDM

Andrew Potterton^{1,2}, Alexander Heifetz^{1,2} and Andrea Townsend-Nicholson^{*1}

*Corresponding Author: Andrea Townsend-Nicholson

Email: a.townsend-nicholson@ucl.ac.uk

¹ Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London, WC1E 6BT, United Kingdom

² Evotec (U.K.) Ltd., 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire, OX14 4RZ, United Kingdom

Synergistic use of GPCR modelling and SDM experiments to understand ligand binding

Summary

There is a substantial amount of historical ligand binding data available from site-directed mutagenesis (SDM) studies of many different GPCR subtypes. This information was generated prior to the wave of GPCR crystal structure, in an effort to understand ligand binding with a view to drug discovery. Concerted efforts to determine the atomic structure of GPCRs have proven extremely successful and there are now more than 80 GPCR crystal structure in the PDB database, many of which have been obtained in the presence of receptor ligands and associated G proteins. These structural data enable the generation of computational model structures for all GPCRs, including those for which crystal structure do not yet exist. The power of these models in designing novel ligands, especially those with improved residence times, and for better understanding receptor function can be enhanced tremendously by combining them synergistically with historic SDM ligand binding data. Here, we describe a protocol by which historic SDM binding data and receptor models may be used together to identify novel key residues for mutagenesis studies.

Key Words

GPCRs; Adenosine receptors; Homology modelling; Ligand binding; Binding kinetics; Receptor; Site-directed mutagenesis

1. Introduction

1.1 Site-directed mutagenesis (SDM) binding studies

Despite their shared seven transmembrane helix structure, GPCRs recognize a wide array of ligands in many different signaling pathways [1]. Ligand specificity stems from sequence variance between receptors, at key residues. The mechanism of specificity is important to understand so that structure-based drug design can achieve high efficacy and avoid off-target side effects. To determine these key

amino acid residues, site-directed mutagenesis (SDM) studies are performed. By comparing binding values for mutant compared to wildtype receptors, the influence of a given residue on ligand binding affinity or kinetics can be determined. Many mutagenesis studies have been conducted in a shotgun approach, but careful targeting of informative mutations for these experiments will enable more precise information about the role of a given residue to be determined.

Extensive mutagenesis studies were performed on the four adenosine receptor subtypes with the aim of determining residues that confer binding specificity for each of the receptor subtypes [2]. In general, these were based on multiple sequence alignments that had been used to identify residues, that differed between the four receptors, with the view that they might contribute to subtype-specific ligand binding. These studies led to the design of a number of different subtype-specific ligands [3]. SDM binding studies are still carried out on the adenosine receptors, mostly performing kinetic binding experiments to rationalize residence time.

1.2 Using GPCR models to elucidate binding

Advances in techniques that stabilize GPCRs, which have many different conformations, have allowed a greater number of these receptors to be crystallized [4]. Further, these stabilization methods have enabled receptors to be co-crystallized, generating structures with agonist bound to the receptor [5]. These methods involve specific thermostabilizing mutations and often include the engineering of a fusion domain between transmembrane helix 5 and 6. If these modifications are restored to the wildtype residues, homology modelling can be used to further increase the number of receptors for which accurate models can be obtained. In homology modelling, the model of the receptor is largely treated as static during drug docking, ignoring ligand flexibility and any conformational changes that could take place upon ligand binding. The loop regions, particularly extracellular loop 2, have been found to be involved in ligand binding [6], causing a problem for traditional GPCR homology modelling as the loop regions tend to be inaccurately predicted. Hierarchical GPCR modelling protocol (HGMP), described in chapter 19, is a more advanced modelling workflow that addresses these problems.

The use of residue engineering and the introduction of non-GPCR sequences to stabilize receptors for improved crystallization means that computational models based on crystal structure are the best means of exploring structure-function relationships for GPCRs. Using accurate models allows the mutagenesis data to be put in the 3D context of the binding site, allowing for indirect interactions to be more easily noted. For the A_{2A} adenosine receptor, viewing mutagenesis studies in the context of a model has enabled the identification of a hydrophobic pocket, that holds the adenine ring of agonists [7]. Historic mutagenesis binding data, therefore, when explored in the context of a computational model can be of great help in understanding ligand binding at the atomic level.

2. Methods

2.1 Models can inform which residues to mutate

The overall process of this workflow is outlined in Fig. 15.1. The first stage of a SDM binding experiment is to plan which residues to mutate, something that requires careful attention in order to maximize the amount of information that can be obtained. Identification of a suitable model, either crystal or homology, should be used as a starting point in planning:

1. Search for a crystal structure for the GPCR to be studied, using the PDB [8].
 - a. If there are multiple entries for the receptor, make a quick table comparing the resolution, the ligand bound, if any, and any crystallization techniques that may have altered the structure of the receptor, including the use of fusion domains and stabilizing mutations.
 - b. If there is a single entry for the receptor of interest, it is important to note these structural features, as they will allow one to assess the reliability of using the structure for SDM binding experiment design. In cases where the receptor has been engineered to look at how the receptor functions, for example by fusion of the receptor with a G_s [9], it may be more suitable to make a homology model for experimental design.

- c. If there are no crystal structures, conduct a literature search to determine whether an appropriate template structure has previously been used for homology modeling, otherwise a template can be generated by homology modeling using the crystal structure of a closely related GPCR:
 - i. Both template and query sequence structure must be numbered according to the Ballesteros-Weinstein numbering scheme [10] to allow for better sequence alignment, which can be achieved with a single command in Molecular Operating Environment, MOE [11].
 - ii. Sequence alignment can be performed in MOE or programs like T-Coffee [12]. MOE is specifically tailored for GPCRs, so that alignment is constrained to align the conserved residues and GPCR motifs, ensuring that a correct fit is likely to be achieved.
 - iii. Homology modelling can also be performed in MOE or using MODELLER [13]. In MOE, side chains are built on a database gathered from a large rotamer library that has been generated by systematic clustering of high-resolution PDB data. The model is then minimized using the OPLS-AA force-field.
 - iv. To check the quality of the model, scoring is performed. GLAS© tests the likeliness of a GPCR to the model structure. If using a new template, however, ProS© should be used instead.
2. It is important to remember that both homology models and crystal structures are static versions of a dynamic system. One therefore needs to explore the conformation space of the ligand to define a binding pocket. Molecular dynamics (MD) simulations can be used to achieve this. An alternative to using MD simulations is to look at multiple models and docking results, as the variability in structures will give a picture of a ligands' conformational space.
3. Search for interactions that are present in the model or MD simulations. Hydrogen bonds can easily be found by predictor tools available in most molecular visualization software

packages, such as PyMOL [14] or UCSF's Chimera [15]. These software packages can also be used to view the hydrophobicity surface of the receptor, which may indicate hydrophobic interactions or buried hydrophilic interactions. Make a list of all these interactions, detailing their nature and in which model/simulation the interaction was seen, (*see Note 1*).

4. Look at the conservation of an interaction, not just in the different models or simulations, but in other related receptors. To look at the equivalent residue in other receptors, the Ballesteros-Weinstein numbering scheme is very useful. This is a numbering system devised for class A GPCRs, allowing a given amino acid to be compared across different GPCRs. In this system, the number before the decimal point indicates what transmembrane helix the amino acid is in, the numbers after it specify the location of this residue compared to the helices conserved residue which, by convention, is defined as .50. For example, the position that corresponds to 5.58, is on the 5th transmembrane helix and is eight positions away from the conserved position [10].
5. Identify the level of structural conservation at a given position, to ascertain whether the interaction is likely to be conserved. It should be noted that the lack of conservation of an interaction may also be valuable when looking at receptor specificity, particularly for models based on nearest neighbors that bind a different endogenous ligand.
6. Once a residue has been selected for mutagenesis, decide which amino acid the residue will be mutated to. The usual choice is alanine, as it prevents that residue from making interactions, is a residue that favors helix formation and is unlikely to cause steric clashes, due to its size. Alanine mutations can, on occasion, prevent stable expression of the receptor in which case other amino acids will need to be considered. In those cases, one can mutate to an amino acid that loses the functional group but retains the relative bulk of the R group. Mutating a position to two or more different amino acids can detail the nature of the interaction. To maximize the information gained from making two or more mutants, one of the mutants should retain the moiety that can make the interaction and the other should lose the interaction but retain the relative bulk of the R group. Mutation to cysteine can be useful for cross-linking experiments.

2.2 Selecting ligands for binding studies using models and SDM binding data

Tailoring the selection of ligands for each mutagenesis study will maximize the information gained from the binding experiments. Four ligands are usually selected for a mutagenesis study that is intended to determine whether a residue interacts with a specific category of ligand. For radioligand binding studies, at least one of the ligands must be available as a radiolabeled ligand to allow for competition binding studies. The following steps will help make this selection of ligand:

1. Look at different interactions made between the receptor and various ligands; to achieve this crystal structures and homology models with ligands docked into their structure must exist (*see Note 2*).
2. From these structures, identify pairs of similar ligands that differ by a single interaction. This is the residue that will be mutated. These ligands may have similar binding affinity values because of their similar structure (*see Note 3*).

2.3 Models and historic SDM binding data add value to the analyses of new SDM binding experiments

After mutagenesis binding studies have been performed, the results should be analyzed in conjunction with models of the receptor to gain a more informed understanding of the nature of the binding and the changes that mutagenesis elicits. One should similarly evaluate any existing historic mutagenesis data with a view to integrating all sources of information needed to put the recently obtained results in context. This will allow for a comprehensive 3D analysis of binding.

1. Gather all mutagenesis binding study data for the receptor of interest. To find previous mutagenesis data for a given receptor, one can use GPCRdb's mutation browser: <http://gpcrdb.org/mutations/> [16]. This will show details of all the mutants that have been made in a specified receptor which can be made into a database. The entries of the database must be checked to indicate whether the mutation was made as part of a binding study or if it was mutated for some other purpose, these latter entries should be removed. The database will allow one to check for historic data for the residue of interest. After the removal of non-binding study data, the ligands used in each study and the fold differences in ligand binding

values to the mutant compared to the wildtype can be added by searching the reference associated with each entry. If no mutants have been made for binding studies with the receptor of interest, another closely related receptor can be used. In these cases, the models of the two receptors should be superimposed, (*see Note 4*), to check for equivalent residues.

2. Map the historic mutagenesis study results to the models of the receptors that have the ligands used in the binding study docked. This can be done by editing the color of residues that have significantly different binding values for a given ligand using a molecular visualization software package, such as PyMOL or Chimera. The current experimental results should also be mapped to the model.
3. Check the position of the residue that has been mutated in comparison with other residues that are colored because of a significant binding difference. This comparison with other residues is not obvious when looking solely at the sequence or the residue numbers because structurally adjacent residues may be on different transmembrane helices, so the model to which the mutagenesis studies were mapped must be used. Significant results located some distance from the predicted binding pocket, as seen from docking or a crystal structure, may indicate a secondary binding site. It is important to look at interactions between the residue of interest and other residues to check for possible indirect interactions with the ligand. The mapped model will indicate whether the residues of indirect interaction play a significant role in binding. The properties of the residues surrounding the mutant residue, in the model, are also important to consider as these may reveal a hydrophobic pocket or a buried hydrophilic interaction.

Fig. 15.2 shows an output of this workflow for the example case of understanding NECA binding to the A_{2A} adenosine receptor. SDM NECA binding data for the A_{2A} receptor was used to make the mapped model [17, 18]. This mapped model can be used to help explain the SDM binding data. Looking at Ser277 and Thr88, in the context of the mapped model, rationalized why these mutants caused loss of binding only to agonists because of their proximity to the ribose ring not present in adenosine receptor antagonists.

2.4 SDM data can be used to select a docking pose

Docking a ligand into a receptor results in many distinct docking poses. Selection of the correct pose from the list of docked positions can be tricky as the only selection criterion given is an arbitrary docking score. Using the following methodology, SDM binding data can give experimental validation for selection of a given pose:

1. Work out the binding pocket of the GPCR. This can be done by looking at the locations of mutagenesis binding studies with significant results for that ligand, or if those data are not available, use any ligand for which data exist. This will give a quick selection criterion that should halve the number of docking results.
2. Check that key interactions, predicted using SDM binding data and the methodology detailed in section 2.3, are possible in each docked pose. Hydrogen bond predictor tools can be useful to help determine this but it is important to remember that the receptor is flexible, which is something that is not accounted for in docking.
3. Make certain that the orientation of the ligand corresponds to most the significant mutagenesis binding studies data; these significant residues can be colored so that they can be easily identified.

3. Notes

1. Distance measurements can be used to increase the conformational space of the ligand, overcoming the issue of having a static model. Residues within a specified distance from the ligand can be selected.
2. If one is docking an agonist, the model must be an active structure. This means that the original crystal structure must have an agonist bound to the receptor.
3. For the adenosine receptors, there are position-substituted ligands that are particularly useful for blocking a potential interaction; these can be compared to the unsubstituted ligand (e.g. CPA and *R*-PIA are both N⁶-substituted agonists).
4. UCSF's Chimera has a 'MatchMaker' feature [15], found under 'Structure Comparison' in the 'Tools' dropdown, which is good at quickly overlaying structures. This finds the best

overall fit between the two receptors however selection can be made to just include specific residues, like those the binding site.

4. References

1. Marinissen MJ, Gutkind JS (2001) G-protein-coupled receptors and signaling networks: Emerging paradigms. *Trends Pharmacol Sci* 22:368–376. doi: 10.1016/S0165-6147(00)01678-3
2. Fredholm BB, IJzerman AP, Jacobson KA, et al (2001) International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. *Pharmacol Rev* 53:527–552. doi: 10.1124/pr.110.003285.1
3. Franchetti P, Cappellacci L, Marchetti S, et al (1998) 2'-C-Methyl Analogues of Selective Adenosine Receptor Agonists: Synthesis and Binding Studies. *J Med Chem* 41:1708–1715.
4. Heydenreich F, Vuckovic Z, Matkovic M, Veprintsev D (2015) Stabilization of G protein-coupled receptors by point mutations. *Front Pharmacol* 6:1–15. doi: 10.3389/fphar.2015.00082
5. Lebon G, Bennett K, Jazayeri A, Tate CG (2011) Thermostabilisation of an agonist-bound conformation of the human adenosine A2A receptor. *J Mol Biol* 409:298–310. doi: 10.1016/j.jmb.2011.03.075
6. Nguyen ATN, Baltos J-A, Thomas T, et al (2016) Extracellular loop 2 of the adenosine A1 receptor has a key role in orthosteric ligand affinity and agonist efficacy. *Mol Pharmacol* 703–714. doi: 10.1124/mol.116.105007
7. Olah ME, Stiles GL (2000) The role of receptor structure in determining adenosine receptor activity. *Pharmacol Ther* 85:55–75. doi: 10.1016/S0163-7258(99)00051-0
8. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. doi: 10.1093/nar/28.1.235
9. Carpenter B, Nehmé R, Warne T, et al (2016) Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature* 536:104–107. doi: 10.1038/nature18966
10. Ballesteros J, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-

- coupled receptors. *Methods Neurosci* 25:366–428. doi: 10.1016/S1043-9471(05)80049-7
11. Chemical Computing Group Inc. (2017) Molecular Operating Environment (MOE), Version 2015.10.
 12. Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217. doi: 10.1006/jmbi.2000.4042
 13. Fiser A, Sali A (2003) MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol* 374:461–491. doi: 10.1016/S0076-6879(03)74020-8
 14. Schrodinger LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.
 15. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi: 10.1002/jcc.20084
 16. Munk C, Isberg V, Mordalski S, et al (2016) GPCRdb: the G protein-coupled receptor database - an introduction. *Br J Pharmacol* 16:2195–2207. doi: 10.1111/bph.13509
 17. Kim J, Wess J, van Rhee AM, et al (1995) Site-directed mutagenesis identifies residues involved in ligand recognition in the human A_{2A} adenosine receptor. *J Biol Chem* 270:13987–97.
 18. Jiang Q, Rhee AM Van, Kim J, et al (1996) Hydrophilic Side Chains in the Third and Seventh Transmembrane Helical Domains of Human A_{2A} Adenosine Receptors Are Required for Ligand Recognition. *Mol Pharmacol* 50:512–521.

Figure Captions

Figure 1. Flow diagram of a simplified version of the methodology detailed in this chapter. Diamond shaped boxes represent decisions that need to be made and the rounded rectangle boxes represents inputs that help carry out a task.

Figure 2. A model of the A_{2A} adenosine receptor with SDM binding data for NECA mapped, using the methodology described in this protocol. Residues, which when mutated have significantly different binding affinity value compared to the wildtype, are colored in pink and are in stick representation. NECA, the ligand, is also shown in stick representation and is tan colored.